**AP Stat- Chapter 7- Correlation Worksheet 7B**

**NEED: Program CORR, group HYPOCORR**

**Ungroup HYPOCORR (you should get lists EXA1, EXA2, EXB1, EXB2, etc.)**

**These lists are hypothetical exam scores, the same ones from Worksheet 7A**

1. Look back at the scatterplots from worksheet 7A to see the graphs for Classes A-F. Now use the lists that you have and compute the correlation for each graph, and fill in the correlation in the same table from before:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Strong** | **Moderate** | **Weak** |
| **Negative** | C -0.9855 | D -0.7198 | F -0.4723 |
| **Positive** | E 0.9888 | A 0.7131 | B 0.4653 |

1. Based on these results, what do you suspect is the largest value that a correlation coefficient can assume?

**-1 < r <1**

1. What do you suspect is the smallest value?

**-1 < r <1**

1. What types of scatterplots have the largest correlation coefficient? Which have the smallest?

**Strong ones**

1. How does the value of the correlation relate to the *direction* of the association?

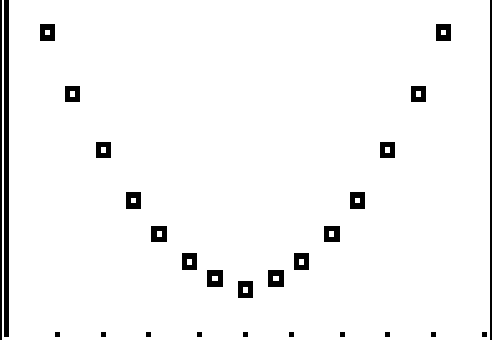
**Negative association 🡪 negative correlation**

1. How does the value of the correlation relate to the *strength* of the association?

**Closer to 1 (or -1) 🡪 stronger plots closer to 0 🡪 weaker plots**

1. Make a scatterplot relating the scores from Exam G (EXG1 = x-list, EXG2 = y-list). Draw the graph. Does there seem to be any relationship between the scores? **Describe** the scatterplot.

**Curved, strong, negative &then positive. There is a strong curved relationship between the scores**



1. Calculate the correlation for Class G. Does its value surprise you? What type of relationship does *r* REALLY measure??

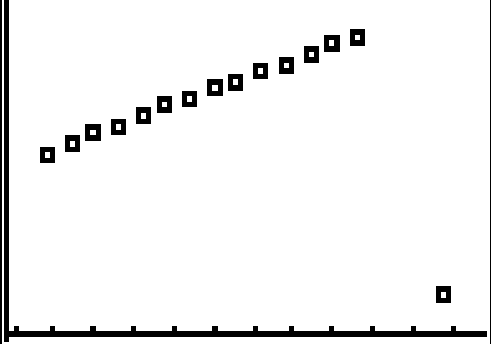
**R = 0. This shows that correlation only measures LINEAR relationships. Not curved. Class G is a very weak LINEAR relationship. But it is a strong CURVED relationship.**

1. Calculate the correlation for Class G but switch the X and Y list. What is it? How does it compare to the original correlation? What does this tell you?

**R = 0. You can switch the X & Y variables, and the correlation remains the same. When calculating correlation, it doesn’t matter which variable you call X or Y.**

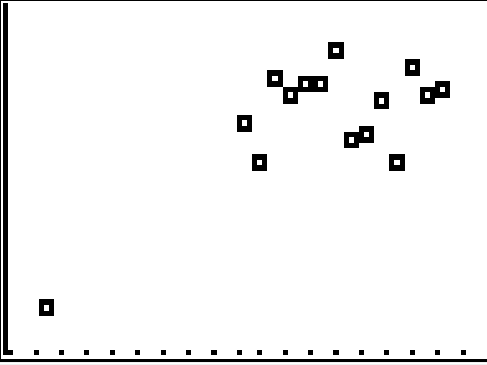
1. Make a scatterplot for Class H. Sketch it below. Do most of the observations seem to follow a linear pattern? Are there any exceptions?

**Most of the observations are linear. Except 1.**



1. Make a scatterplot for Class I. Do most of the observations seem to be scattered? Are there any exceptions?

**Yes, scattered except for 1.**



1. Calculate the correlation for Classes H and I. Write them both down. Do either of these correlations surprise you? Why?

**Class H: r = 0.0365 I think this should be closer to 1 since the data is pretty linear**

**Class I: r = 0.7046 I think this should be closer to 0 since the data is non linear**

1. Remove the outlier for Class H (be sure to remove the coordinates from both the X and Y lists). Recalculate the correlation. How has it changed?

**r = 0.997 It increased greatly, since the plot is now almost perfectly linear. Outliers can have a**

**negative effect on the correlation if they make the plot less linear.**

1. Remove the outlier for Class I. Recalculate the correlation. How has it changed?

**r = 0.1304 It dropped significantly, since the plot is now basically scattered. Outliers can help increase**

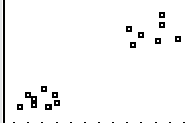
**the correlation sometimes, if they make the plot more linear.**

1. Based on your analysis of Classes H and I, would you say that the correlation coefficient is *resistant* measure of association? Or *non-resistant*?

**Correlation is NON-RESISTENT. It is affected by outliers.**

1. Make a scatterplot for Class J. Describe what the plot reveals about the relationship between exam scores (describe the plot).

**Two clusters, positive association**



1. Calculate the correlation for Class J. Does this value surprise you? How?

**r = 0.9544. This is higher than expected.**